THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 288, NO. 9, pp. 6351–6362, March 1, 2013

Structural Basis of the Versatile DNA Recognition Ability of the Methyl-CpG Binding Domain of Methyl-CpG Binding Domain Protein 4*

Received for publication, October 26, 2012, and in revised form, January 8, 2013 Published, JBC Papers in Press, January 10, 2013, DOI 10.1074/jbc.M112.431098

Junji Otani[‡], Kyohei Arita[‡], Tsuyoshi Kato[‡], Mariko Kinoshita[‡], Hironobu Kimura[§], Isao Suetake[§], Shoji Tajima[§], Mariko Ariyoshi[¶‖1], and Masahiro Shirakawa[‡‡‡2]**

*From the [‡]Graduate School of Engineering, Kyoto University, Kyoto 615-8510, the [§]Institute for Protein Research, Suita 565-0871, the [¶]Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto 615-8501, the JST, [‖]PREST and **CREST, Saitama 332-0012, and the [‡‡]Yokohama Institute, RIKEN, Kanagawa 230-0045, Japan*

Background: Methyl-CpG binding domain 4 (MBD4) is a DNA glycosylase that excises mismatched bases generated in methylated CpG sequences.
Results: We report the biochemical and structural properties of the methyl-CpG binding domain of MBD4 (MBD_{MBD4}).
Conclusion: MBD_{MBD4} recognizes a wide range of 5-methylcytosine modifications via an extensive hydration network.
Significance: This study provides new insight into the structural mechanism of the broad base recognition that is unique to MBD_{MBD4}.

The methyl-CpG binding domain (MBD) protein MBD4 participates in DNA repair as a glycosylase that excises mismatched thymine bases in CpG sites and also functions in transcriptional repression. Unlike other MBD proteins, MBD4 recognizes not only methylated CpG dinucleotides (^{5m}CG/^{5m}CG) but also T/G mismatched sites generated by spontaneous deamination of 5-methylcytosine (^{5m}CG/TG). The glycosylase activity of MBD4 is also implicated in active DNA demethylation initiated by the deaminase-catalyzed conversion of 5-methylcytosine to thymine. Here, we report the crystal structures of the MBD of MBD4 (MBD_{MBD4}) complexed with ^{5m}CG/^{5m}CG and ^{5m}CG/TG. The crystal structures show that the DNA interface of MBD4 has flexible structural features and harbors an extensive water network that supports its dual base specificities. Combined with the results of biochemical analyses, the crystal structure of MBD4 bound to 5-hydroxymethylcytosine further demonstrates that MBD_{MBD4} is able to recognize a wide range of 5-methylcytosine modifications through the unique water network. The versatile base recognition ability of MBD_{MBD4} implies multifunctional roles for MBD4 in the regulation of dynamic DNA methylation patterns coupled with deamination and/or oxidation of 5-methylcytosine.

DNA methylation is the most prominent epigenetic modification in higher eukaryotic genomes (1, 2). In mammals, DNA methylation mainly occurs at the C5 position of symmetrically arranged cytosines in CpG dinucleotides, and plays essential roles in various cellular events such as gene repression, imprinting, X-chromosome inactivation, suppression of repetitive genomic elements, and carcinogenesis (3). Recent studies have shown that DNA methylation can be actively reversed and that its pattern is dynamically altered in mammalian cells (4–7). Although the underlying molecular mechanism is not fully understood, active DNA demethylation has been proposed to involve further oxidation or deamination of 5-methylcytosine (^{5m}C)[3] followed by base excision repair (6–13). Successive oxidation of ^{5m}C to 5-hydroxymethylcytosine (^{hm}C), 5-formylcytosine (^{fo}C), and 5-carboxylcytosine (^{ca}C) is catalyzed by TET proteins and, has attracted much attention as a crucial process in DNA demethylation. Furthermore, demethylation pathways are thought to involve the spontaneous or enzymatic deamination of ^{5m}C or ^{hm}C and subsequent base excision repair of the mismatched thymine or 5-hydroxymethyluracil (^{hm}U) base (6, 9, 11–13). Therefore, precise interpretation and regulation of the modification status of ^{5m}C are required for various epigenetic events in cells.

MBD (methyl-CpG binding domain) proteins are archetypal mediators of DNA methylation marks. They recognize methyl-CpG sites (^{5m}CG/^{5m}CG) through a conserved MBD and recruit transcriptional repressors or chromatin modifiers to these sites (14). One of the MBD family proteins, MBD4 contains a C-terminal DNA glycosylase domain in addition to an N-terminal MBD domain. MBD4 is involved in DNA mismatch repair as a T/G or U/G mismatch glycosylase and also in transcriptional repression via its recruitment of Sin3A and HDAC1 (15, 16).

* This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology and the Japan Science and Technology Agency (MEXT) (to M. S.) and by JST, PRESTO (to M. A.) and the Global COE Program "International Center for Integrated Research and Advanced Education in Materials Science" (No. B-09) of MEXT of Japan, administered by the Japan Society for the Promotion of Science.
The atomic coordinates and structure factors (codes 3VXV, 3VXX, 3VYB, and 3VYQ) have been deposited in the Protein Data Bank (http://wwpdb.org/).
[1] To whom correspondence may be addressed: Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto 615-8501, Japan. Tel.: 81-75-383-2536; Fax: 81-75-383-2541; E-mail: ariyoshi@moleng.kyoto-u.ac.jp.
[2] To whom correspondence may be addressed: Graduate School of Engineering, Kyoto University, Kyoto 615-8510, Japan. Tel.: 81-75-383-2535; Fax: 81-75-383-2541; E-mail: shirakawa@moleng.kyoto-u.ac.jp.

[3] The abbreviations used are: ^{5m}C, 5-methylcytosine; ^{hm}C, 5-hydroxymethylcytosine; ^{fo}C, 5-formylcytosine; ^{ca}C, 5-carboxylcytosine; ^{hm}U, 5-hydroxymethyluracil; TDG, thymine DNA glycosylase; MBD, methyl-CpG binding domain; ITC, isothermal titration calorimetry; PDB, Protein Data Bank.

The glycosylase activity of MBD4 specifically excises a mismatched thymine or ^{hm}U base generated by the deamination of ^{5m}C or ^{hm}C in a CpG site; thus, MBD4 is thought to participate in both DNA repair in the context of CpG and DNA demethylation (14). The functional importance of MBD4 in maintaining genomic integrity has been demonstrated by an increased frequency of C to T transitions at CpG sites in MBD4$^{-/-}$ mice (17) and the finding that frequent MBD4 mutations in various human carcinomas are characterized by microsatellite instability (18). Moreover, MBD4 contributes to the stimuli-dependent active DNA demethylation of specific genomic loci together with thymine DNA glycosylase (TDG) (19).

Previous structural studies of MBD1, MBD2, and MeCP2 demonstrated how MBDs recognize only ^{5m}CG/^{5m}CG sites (20–22). However, in addition to the fully methylated CpG, the MBD domain of MBD4 binds to T/G mismatched base pairs that result from asymmetrical ^{5m}C deamination of ^{5m}CG/^{5m}CG dinucleotides (16). Recent structural and biochemical studies of the glycosylase domain of MBD4 suggest that the specificity of full-length MBD4 for ^{5m}CG/TG is provided by MBD$_{MBD4}$ (23–25). The glycosylase domain recognizes the mismatched thymine or ^{hm}U base but not the adjacent ^{5m}C/G base pair. Thus, the recognition of methylated DNA by MBD$_{MBD4}$ appears to be indispensable for the multifunctional roles of MBD4 in the regulation and maintenance of DNA methylation patterns.

Here, we present the crystal structures of the MBD of MBD4 (MBD$_{MBD4}$) in complex with a DNA fragment containing the ^{5m}CG/^{5m}CG site or its deamination product, ^{5m}CG/TG. The structures reveal the unique flexible DNA interface of MBD$_{MBD4}$ accompanied by an extensive water network. Our structural and biochemical data demonstrate that, in addition to ^{5m}CG/^{5m}CG and ^{5m}CG/TG, the DNA interface of MBD$_{MBD4}$ is able to accommodate ^{hm}C and its further oxidation or deamination products. We also determined the crystal structure of MBD$_{MBD4}$ bound to a methylated CpG site containing ^{hm}C (^{5m}CG/^{hm}CG) and found that the water network at the DNA interface of MBD$_{MBD4}$ can be finely tuned to accommodate various modified pyrimidine rings. Our structural and biochemical studies indicate the molecular basis of the broad base recognition ability of MBD$_{MBD4}$, which underlies DNA methylation and gene regulation involving MBD4.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—A DNA fragment encoding MBD$_{MBD4}$ (residues 69–136) was amplified by PCR and cloned into the bacterial expression vector pGEX4T-3 (GE Healthcare Biosciences), which was engineered for the expression of recombinant proteins with an N-terminal tandem fusion tag of glutathione *S*-transferase (GST) and small ubiquitin-like modifier-1 (SUMO-1). The GST-SUMO-1-MBD$_{MBD4}$ fusion was overexpressed in *Escherichia coli* strain BL21(DE3). Cells were grown at 37 °C in Luria-Bertani (LB) medium containing 50 μg/ml of ampicillin, to an optical density of 0.5–0.6 at 660 nm, and then induced with 0.2 mM isopropyl β-D-thiogalactoside for 15 h at 18 °C. Cells were harvested by centrifugation, and lysed by sonication in 50 mM Tris-HCl, pH 8.0, buffer containing 300 mM NaCl, 1 mM dithiothreitol (DTT), 5% glycerol, 0.1% Triton X-100, and 1 mM phenylmethylsulfonyl fluoride. The clarified lysate was loaded onto glutathione-Sepharose 4 Fast Flow beads (GE Healthcare). GST-SUMO-1-fused MBD$_{MBD4}$ was eluted from the beads with elution buffer containing 10 mM glutathione. The tag-free MBD$_{MBD4}$ was prepared by SENP2 protease treatment, and was further purified by sequential column chromatography steps using HiTrap Heparin HP and HiLoad 16/60 Superdex 75 columns (GE Healthcare). Purified protein in the final elution buffer containing 10 mM Hepes-NaOH, pH 7.4, 150 mM NaCl, and 2 mM DTT was concentrated using an Amicon Ultra 3,000 cut-off membrane concentrator (Millipore). To introduce selenomethionine, Leu-116 of MBD4 was substituted with methionine. The selenomethionine containing MBD$_{MBD4}$ was expressed in modified M9 medium (26). Purification of the selenomethionine-labeled L116M mutant was performed following the same procedure as that for the native protein.

Crystallization, Data Collection, and Structure Determination—MBD$_{MBD4}$ at a concentration of 200–800 μM was mixed with each DNA fragment at a 1:1 molar ratio. Crystals of MBD$_{MBD4}$ were obtained by a vapor diffusion method at 20 °C using PEG 10,000 or PEG 1500 as the precipitant. Details of crystallization conditions are listed in Table 1. MBD$_{MBD4}$ bound to 14- and 11-bp oligomers containing ^{5m}CG/TG were crystallized in orthorhombic $C222_1$ and triclinic $P1$ forms, respectively. In the orthorhombic form, a complex of one protein and DNA is contained in an asymmetric unit, whereas the triclinic form comprises two protein molecules and one DNA oligomer. The complex of MBD$_{MBD4}$ with 14-bp oligomer containing ^{5m}CG/^{5m}CG or ^{5m}CG/^{hm}CG was crystallized in a $C222_1$ form. All crystals were flash frozen at 100 K in cryoprotectant containing 20% ethylene glycol. X-ray diffraction data sets were collected at a wavelength of 1.0000 Å on beamlines BL-5A, BL-17A, NE3A, and NW12 at Photon Factory (Tsukuba, Japan) and beamline BL-38 at SPring8 (Harima, Japan), and were processed with the program HKL2000 (27). The phases of the selenomethionine derivative MBD$_{MBD4}$ L116M complexed with the 14-bp ^{5m}CG/TG fragment were determined by the single wavelength anomalous dispersion method using the programs SOLVE and RESOLVE (28, 29). The initial model was built using the COOT program (30) and was refined against the native data using the PHENIX suite (31), thus yielding a crystallographic R factor of 18.8% and a free R factor of 22.4% to 2.0 Å. The triclinic form structure of the MBD$_{MBD4}$-^{5m}CG/TG complex and the structures of the MBD$_{MBD4}$-^{5m}CG/^{5m}CG and MBD$_{MBD4}$-^{5m}CG/^{hm}CG complexes were solved by a molecular replacement method using the orthorhombic form structure of MBD$_{MBD4}$-^{5m}CG/TG as the search model. The stereochemical quality of the final models was assessed using MolProbity (32). The sequence information of DNA fragments used for crystallization is summarized in Table 2. The crystallographic data, data collection statistics, and refinement statistics are summarized in Table 1. All structural figures were produced using PyMOL (43).

DNA Binding Assays—Isothermal titration calorimetry (ITC) measurements were performed on an iTC200 microcalorimeter (MicroCal, USA) at 25 °C. The protein solution was dialyzed to the ITC measurement buffer of 25 mM Hepes-NaOH, pH 7.4, containing 100 mM NaCl and 0.1 mM Tris(2-carboxy-

TABLE 1
Crystallographic data and refinement statistics

Crystal	14 bp				11 bp
	^{5m}CG/TG (SeMet)	^{5m}CG/TG	^{5m}CG/^{5m}CG	^{5m}CG/^{hm}CG	^{5m}CT/TG
Crystallization condition	12% PEG10,000 0.1 M Na acetate, pH 4.4 0.2 M NaCl	12% PEG10,000 0.1 M Na acetate, pH 4.4 0.2 M NaCl	7% PEG10,000 0.1 M Na acetate, pH 4.4 0.2 M NaCl	8% PEG1,500 0.1 M Na acetate, pH 3.9 0.1 M Na cacodylate, pH 5.4	2 mM Zn acetate 0.1 M Na cacodylate
X-ray source	PF-NE3A	PF BL-5A	PF-NW12	PF-BL17A	SPring8-BL38B1
Wavelength (Å)	0.97923 (peak)	1.0000	1.0000	1.0000	1.0000
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$P1$
Unit cell parameters (Å, °)	$a = 88.752$ $b = 97.588$ $c = 54.959$ $\alpha = \beta = \gamma = 90$	$a = 89.074$ $b = 94.989$ $c = 54.738$ $\alpha = \beta = \gamma = 90$	$a = 89.182$ $b = 93.829$ $c = 55.357$ $\alpha = \beta = \gamma = 90$	$a = 88.693$ $b = 97.758$ $c = 55.725$ $\alpha = \beta = \gamma = 90$	$a = 30.324$ $b = 33.781$ $c = 60.035$ $\alpha = 75.317$ $\beta = 77.743$ $\gamma = 87.352$
Resolution range (Å)[a]	50–2.7 (2.8–2.7)	50–2.0 (2.07–2.0)	50–2.2 (2.28–2.2)	50–2.19 (2.27–2.19)	50–2.5 (2.59–2.5)
Total observations	38,796	114,920	86,212	74,818	20,848
Unique reflections[a]	6,051 (363)	16,062 (1591)	12,048 (1192)	11,517 (701)	7,225 (531)
Multiplicity[a]	6.4 (5.6)	7.2 (7.3)	7.2 (7.3)	6.5 (4.1)	2.9 (1.7)
R_{merge}[a,b]	0.097 (0.308)	0.031 (0.365)	0.078 (0.477)	0.052 (0.317)	0.058 (0.125)
Completeness (%)[a]	88.6 (54.0)	99.7 (100)	99.6 (99.9)	89.6 (54.7)	92.8 (67.4)
$\langle I/\sigma(II)\rangle$[a]	12.7 (5.2)	18.2 (4.3)	12.4 (4.5)	14.7 (4.1)	16.3 (5.4)
Refinement	Resolution range (Å)	32.5–2.00	28.3–2.20	34.7–2.40	33.0–2.53
	R_{work} (%)[c]	18.8	19.6	19.06	19.34
	R_{free} (%)[c]	22.4	21.4	21.99	23.65
Root mean square deviations	Bond length (Å)	0.018	0.006	0.004	0.007
	Bond angle (°)	2.071	1.193	1.008	1.161
Ramachandran plot	Favored (%)	100	100	98.36	99.18
	Allowed (%)	100	100	100	100

[a] Numbers in parentheses are the values for the highest resolution shell of each data set.
[b] $R_{merge} = \Sigma_h \Sigma_i |I(h)i - \langle I(h)\rangle| / \Sigma_h \Sigma_i I(h)_i$, where $I(h)$ is the intensity of reflection h, Σ_h is the sum of all measured reflections and Σ_i is the sum of i measurements of reflection.
[c] R_{work} and $R_{free} = (\Sigma hkl \, ||F_o| - |F_c||) / \Sigma hkl \, |F_o|$, where the free reflections (5% of the total used) were held aside for R_{free} throughout refinement.

ethyl)phosphine. Each annealed DNA duplex was dried and dissolved in ITC buffer. The DNA solution (10–20 μM) in a calorimetric cell was titrated with a 100–400 μM protein solution. Binding constants were calculated by fitting the data using the ITC data analysis module of Origin 7.0 (OriginLab). Competitive binding assays were also performed in the ITC buffer. The upper strand of the 14-bp ^{5m}CG/^{5m}CG DNA fragment was radioisotope labeled at the 5′ end with T4 polynucleotide kinase (TOYOBO, Japan) and [γ-^{32}P]ATP (Muromachi Kagaku, Tokyo). The labeled strand was then mixed with a 1.2-fold amount of the complementary strand and annealed. The radioisotope-labeled ^{5m}CG/^{5m}CG fragment and MBD protein were mixed at concentrations of 1 and 3 μM, respectively. Subsequently, 0, 1, 2, or 4 μM nonlabeled competitor DNA fragment was added, and the solution was incubated for 30 min at 4 °C and analyzed using native gel electrophoresis. The DNA bands were visualized with a Fuji BAS-2000 phosphorimager. The DNA content of each band was quantified from the gel band density as a relative amount compared with total input DNA. A series of relative values were normalized against the control lane and plotted against the amount of competitor DNA. Curves for each experiment were fitted by a nonlinear, least square method using Morrison's equation (33).

Glycosylase Assays—Glycosylase assays were performed in a 10-μl reaction mixture containing 10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, and 0.1 mg/ml of BSA. The synthetic oligonucleotides, GGCTAAATACCTGGGCT**X**GAAGTGAACTGATTGCC, where **X** indicates T, ^{hm}U, ^{ca}C, ^{5m}C, ^{hm}C, or ^{fo}C, was labeled by T4 polynucleotide kinase and [γ-^{32}P]ATP, and annealed with complementary strand containing ^{5m}C at the central CpG site. Each of the ^{32}P radioisotope-labeled DNA duplexes at 40 or 400 nM were incubated with 200 or 2000 nM human TDG or mouse MBD4 at 37 °C for 1 h. Reactions were terminated by the addition of 10 μl of a reaction stop solution containing 0.2 M NaOH and 20 mM EDTA followed by incubation at 95 °C for 10 min. For the reaction using the ^{fo}C-containing oligonucleotide, the incubation was carried out at 70 °C for 5 min to avoid the digestion under the alkaline conditions regardless of enzymatic activity. After addition of 60 μl of 10 M urea, 20 μl of each sample was subjected to electrophoresis in a 9 M urea, 20% PAGE and visualized with a Fuji BAS-2000 phosphorimager.

Accession Codes—Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession codes 3VXV, 3VXX, 3VYB, and 3VYQ.

RESULTS

Dual Binding Specificity of MBD$_{MBD4}$ for ^{5m}CG/^{5m}CG and ^{5m}CG/TG Sites—The DNA binding properties of mouse MBD$_{MBD4}$ were examined quantitatively by ITC measurements using 14-bp double-stranded DNA oligomers containing a single CpG site in various modification or mismatch states (Tables 2 and 3). In agreement with previous reports (16), MBD$_{MBD4}$ was tightly bound to ^{5m}CG/TG with a dissociation constant (K_D) of 98.8 nM, but did not interact with nonmethylated CG/CG or hemimethylated ^{5m}CG/CG. The affinity of MBD$_{MBD4}$ for the ^{5m}CG/^{5m}CG site (K_D, 97.5 nM) was comparable with that for the ^{5m}CG/TG site. However, MBD$_{MBD1}$ exhibited a 5-fold greater affinity for the ^{5m}CG/^{5m}CG site (K_D, 72.5 nM) over the ^{5m}CG/TG site (K_D, 458 nM). The binding

TABLE 2
DNA sequences used in crystallization and binding assays

	Lower strand	Upper strand
11 bp ^{5m}CG/TG	TCAC **TG** GATGT	ACATC ^{5m}**CG** GTGA
14 bp ^{5m}CG/TG	GTC **TG** GTAGTGACT	GTCACTAC ^{5m}**CG** GACA
^{5m}CG/^{hm}UG	GTC ^{hm}**UG** GTAGTGACT	GTCACTAC ^{5m}**CG** GACA
^{5m}CG/^{ca}CG	GTC ^{ca}**CG** GTAGTGACT	GTCACTAC ^{5m}**CG** GACA
^{5m}CG/^{5m}CG	GTC ^{5m}**CG** GTAGTGACT	GTCACTAC ^{5m}**CG** GACA
^{5m}CG/^{hm}CG	GTC ^{hm}**CG** GTAGTGACT	GTCACTAC ^{5m}**CG** GACA
^{5m}CG/^{fo}CG	GTC ^{fo}**CG** GTAGTGACT	GTCACTAC ^{5m}**CG** GACA
^{hm}CG/^{hm}CG	GTC ^{hm}**CG** GTAGTGACT	GTCACTAC ^{hm}**CG** GACA
CG/CG	GTC **CG** GTAGTGACT	GTCACTAC **CG** GACA
^{5m}CG/CG	GTC **CG** GTAGTGACT	GTCACTAC ^{5m}**CG** GACA
CG/TG	GTC **TG** GTAGTGACT	GTCACTAC **CG** GACA

TABLE 3
Thermodynamic parameters obtained by ITC experiments

DNA	MBD4	MBD1
^{5m}CG/^{5m}CG		
K_D (nM)	97.5 ± 76	72.5 ± 11
ΔH (kJ/mol)	−7.69 ± 0.93	−54.7 ± 5.1
$-T\Delta S$ (kJ/mol)	−32.8 ± 2.5	13.9 ± 5.3
^{5m}CG/TG		
K_D (nM)	98.8 ± 42	458 ± 92
ΔH (kJ/mol)	−18.5 ± 1.7	−46.7 ± 1.3
$-T\Delta S$ (kJ/mol)	−21.7 ± 0.71	10.4 ± 0.86
^{5m}CG/^{hm}CG		
K_D (nM)	162 ± 58	1040 ± 422
ΔH (kJ/mol)	−5.0 ± 1.3	−49.5 ± 1.2
$-T\Delta S$ (kJ/mol)	−33.9 ± 2.3	15.1 ± 0.1
CG/TG		
K_D (nM)	213 ± 58	4025 ± 45
ΔH (kJ/mol)	−11.5 ± 0.3	−43.2 ± 1.0
$-T\Delta S$ (kJ/mol)	−26.7 ± 1.0	12.3 ± 1.1
^{hm}UG/^{5m}CG		
K_D (nM)	287 ± 78	
ΔH (kJ/mol)	−15.0 ± 1.3	Not performed
$-T\Delta S$ (kJ/mol)	−22.3 ± 1.8	
^{5m}CG/CG		
K_D (nM)		3080 ± 830
ΔH (kJ/mol)	Not detected	−36.8 ± 0.9
$-T\Delta S$ (kJ/mol)		5.2 ± 1.6
CG/CG		
K_D (nM)		33,000 ± 6,900
ΔH (kJ/mol)	Not detected	−15.3 ± 1.9
$-T\Delta S$ (kJ/mol)		−10.3 ± 2.3

specificities of MBD$_{MBD4}$ and MBD$_{MBD1}$ were also assessed by a competitive electrophoretic mobility shift assay (EMSA) in which the ^{32}P-labeled ^{5m}CG/^{5m}CG oligomer bound to MBD competed with nonlabeled fragments. Both nonlabeled ^{5m}CG/TG and ^{5m}CG/^{5m}CG fragments efficiently competed with the ^{5m}CG/^{5m}CG oligomer for binding to MBD$_{MBD4}$, whereas the nonlabeled ^{5m}CG/TG fragment did not abrogate the interaction between MBD$_{MBD1}$ and the ^{5m}CG/^{5m}CG site (Fig. 1). Thus, MBD$_{MBD4}$ is characterized as a unique MBD family protein based on its dual DNA binding ability, although the key residues for recognition of methylated CpG sites are almost completely conserved among MBD$_{MBD1}$, MBD$_{MBD2}$, and MBD$_{MeCP2}$ (Fig. 2*A*).

Crystal Structures of MBD$_{MBD4}$ Complexed with Methylated CpG and Its Deamination Product—To understand the structural basis of the unique DNA binding properties of MBD4, we solved the crystal structure of MBD$_{MBD4}$ in complex with DNA fragments containing ^{5m}CG/TG or ^{5m}CG/^{5m}CG sequences (Table 2). The crystal structures of the MBD$_{MBD4}$-^{5m}CG/TG complex were determined in orthorhombinc $C222_1$ and triclinic $P1$ forms (Fig. 2, *B* and *C*, Table 1). In the $C222_1$ form, the C-terminal parts of MBD$_{MBD4}$ (residues 121–136) were exchanged between 2-fold symmetry-related molecules, resulting in a swapped dimer linked through a disulfide bond (Fig. 2*B*). However, dimer formation was not observed in our gel-filtration experiments (data not shown); therefore, the swapped dimer is interpreted as a crystallographic artifact. The $C222_1$ form structure shares the core folding and the DNA binding interface with the $P1$ form despite the C-terminal swapping. In this report, the recognition of ^{5m}CG/TG by MBD$_{MBD4}$ is discussed based on the higher resolution structure of the $C222_1$ form. The DNA strand containing a mismatched T is hereafter referred to as the "lower strand," whereas the other strand is termed the "upper strand" (Fig. 2*D*).

Similar to other MBD family proteins, MBD$_{MBD4}$ has an overall fold consisting of one α-helix (α1) and three β-strands (β1–3) (Fig. 2*D*) (20–22). The overall structures of MBD$_{MBD4}$ and MBD$_{MeCP2}$ are well superimposed with root mean square deviations of 1.59 Å for 54 equivalent Cα atoms. The T/G mismatch DNA fragment bound to MBD$_{MBD4}$ adopts the canonical B-form conformation in both crystal structures. The ^{5m}CG/TG site is recognized by MBD$_{MBD4}$ in a major groove as previously observed in other MBD-^{5m}CG/^{5m}CG complexes (Fig. 2*D*) (20–22). Phosphate backbone recognition is also conserved among MBD family members. The positive end of the helix dipole from the α1 helix is placed in the major groove and capped by a phosphate group from the DNA backbone. Residues 85–89 in the L1 loop, which connect β1 and β2, also assist in holding the phosphate backbone via making extensive electrostatic contacts (Fig. 2*D*).

In the triclinic crystal structure, one of the MBD$_{MBD4}$ molecules in an asymmetric unit binds to the ^{5m}CG/TG site in a conserved manner, whereas the other protein molecule interacts with a joint region between two neighboring DNA fragments that are continuously linked through base stacking interactions (Fig. 2*C*). As described below, the latter protein-DNA interaction suggests a possible mode of nonspecific DNA binding for MBD4.

MBD$_{MBD4}$ complexed with ^{5m}CG/^{5m}CG was crystallized in the $C222_1$ form, and its structure was determined at 2.2 Å resolution (Table 1). The overall structure of the ^{5m}CG/^{5m}CG complex is almost identical to that observed in the orthorhombic crystal of the ^{5m}CG/TG complex (root mean square deviation: 0.16 Å for 57 equivalent Cα atoms).

MBD$_{MBD4}$ Recognizes the ^{5m}CG/^{5m}CG Sequence via Conserved Arginine Fingers—The overall ^{5m}CG/^{5m}CG recognition mode of MBD$_{MBD4}$ is essentially analogous to that of MBD$_{MeCP2}$, MBD$_{MBD1}$, and MBD$_{MBD2}$. Arg-84 and Arg-106, which are completely conserved in the MBD family (Fig. 2*A*), recognize symmetrically arranged guanine bases in the ^{5m}CG/^{5m}CG sequence in a manner similar to that of other MBD proteins (Fig. 2*D*). The Arg-84 and Arg-106 residues are hereafter termed Arg finger-1 and -2, respectively. A guanidino group of Arg finger-1 donates hydrogen bonds to the O6 and N7 atoms of the guanine base in the lower DNA strand, whereas Arg finger-2 recognizes the guanine base in the upper strand via an analogous hydrogen bonding pattern (Fig. 3, *A* and *B*). The aliphatic side chains of each arginine finger make van der Waals contacts with the 5-methyl group of the ^{5m}C base adjacent to its

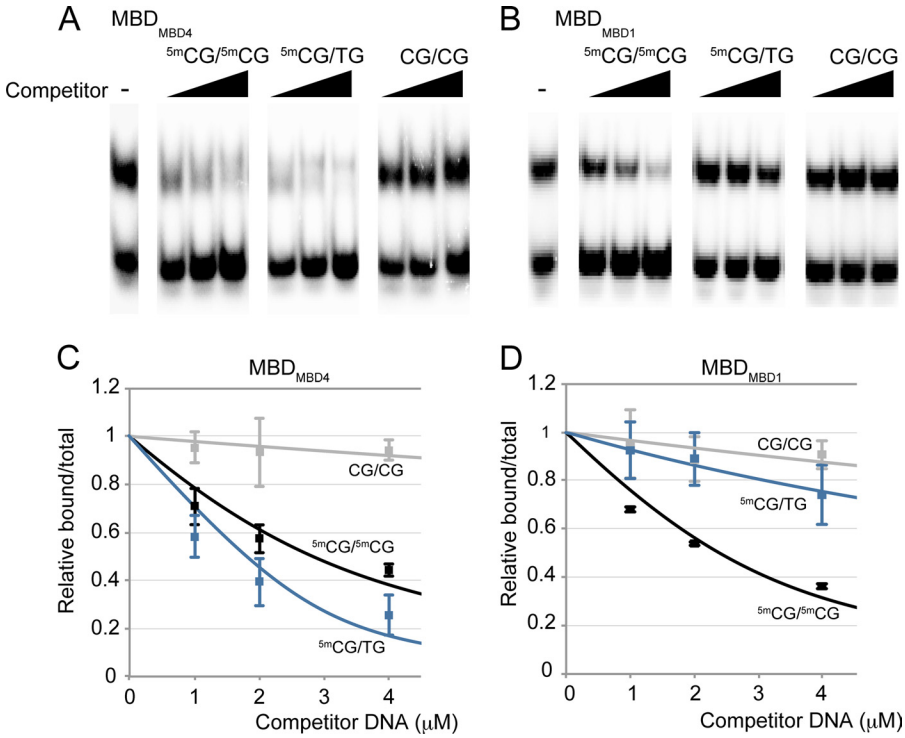

FIGURE 1. **Binding of MBD$_{MBD4}$ to ^{5m}CG/^{5m}CG and ^{5m}CG/TG.** *A* and *B*, representative autoradiographic images of competition assays for MBD$_{MBD4}$ (*A*) and MBD$_{MBD1}$ (*B*). Each competitor sequence is indicated *above* the lanes. The *left panel* in each section shows a control binding experiment in the absence of competitor DNA. *C* and *D*, each band was quantified relative to the total input DNA from the gel band density. The relative values were normalized against the control lane, and plotted against the amounts of competitor DNA. Each data point represents an average of three independent experiments using MBD$_{MBD4}$ (*C*) or MBD$_{MBD1}$ (*D*). Each *solid line* shows a curve fitted using Morrison's equation. An ~1.5-fold excess of the nonlabeled ^{5m}CG/^{5m}CG fragment was required to obtain the same competitive effect as with nonlabeled ^{5m}CG/TG, indicating that MBD$_{MBD4}$ has a similar affinity for ^{5m}CG/TG and ^{5m}CG/^{5m}CG sites.

interacting guanine (Fig. 3, *C* and *D*). Additionally, the main chain carbonyl group of Arg finger-2 forms a CHO hydrogen bond (3.8 Å) with the 5-methyl group of the ^{5m}C base in the upper strand (Fig. 3*D*).

In MBD$_{MeCP2}$, the positions of both Arg fingers are stabilized through interactions with the conserved acidic residues (Fig. 3, *E* and *F*) (21). Similarly, the orientation of Arg finger-1 of MBD$_{MBD4}$ is defined by its intramolecular interaction with a conserved acidic residue, Asp-94. The side chain carboxyl group of Asp-94 forms salt bridges with the guanidino group of Arg finger-1, resulting in an arginine side chain conformation suitable for recognition of the ^{5m}CG sequence (Fig. 3*A*). Asp-94 also forms a CHO hydrogen bond (3.9 Å) with the 5-methyl group of the ^{5m}C base in the lower strand (Fig. 3*C*). In contrast, Arg finger-2 lacks such an intramolecular lock because Glu-137 in MeCP2 is replaced by Ser-110 in MBD4 (Fig. 2*A*). Nevertheless, the position of Arg finger-2 bound to ^{5m}CG/^{5m}CG shows good superimposition with that in the MBD$_{MeCP2}$-^{5m}CG/^{5m}CG complex (Fig. 3, *B* and *F*).

The Water Network in the MBD$_{MBD4}$-DNA Interface—The most significant structural difference between MBD$_{MBD4}$ and other MBD proteins is the orientation of the conserved tyrosine residue, Tyr-96, located on the DNA binding surface (Fig. 3*B*). The corresponding tyrosine residues of MBD$_{MeCP2}$, MBD$_{MBD1}$, and MBD$_{MBD2}$ are oriented toward the ^{5m}C base in the lower strand through hydrophobic interactions with the aliphatic side chains of their surrounding residues (20–22). In the crystal structure of the MBD$_{MeCP2}$-DNA complex, the side chain of the corresponding residue, Tyr-123, recognizes the ^{5m}C base via two water-mediated interactions (Fig. 3*F*) (21). Previous mutational analysis of MBD1 and MeCP2 suggested that the conserved Tyr residue is critical for DNA binding (20, 21). The side chain of Tyr-96 in MBD$_{MBD4}$ is flipped out of the DNA interface and makes water-mediated interactions with the phosphate backbone of the lower DNA strand (Fig. 3*B*). The aromatic side chain is stabilized by a stacking interaction with the compact hydrophobic side chain of Val-80 (Fig. 4). Notably, despite the absence of the tyrosine hydroxyl group at the common position, the MBD$_{MBD4}$-DNA interface retains the hydration water molecules involved in the recognition of the lower strand ^{5m}C (Fig. 3, *B* and *F*). The water molecules, W1, W2, and W3, form van der Waals interactions with the 5-methyl group of the lower strand ^{5m}C in a similar manner to that observed in the MBD$_{MeCP2}$-DNA complex. Coordination of the three other water molecules (W4, W5, and W6 in the MBD$_{MBD4}$-DNA complex) surrounding the upper strand ^{5m}C base is also conserved in the MBD$_{MBD4}$-DNA and MBD$_{MeCP2}$-DNA complexes (Fig. 3, *A* and *E*). MBD$_{MBD4}$ and MBD$_{MeCP2}$ share the recognition scheme for the upper strand ^{5m}C base involving a water molecule (W4 in the MBD4-^{5m}CG/^{5m}CG complex or W4′ in the MeCP2-DNA complex) that bridges the N4 atom of the base with the carboxyl group of the conserved Asp residue.

In the vacant space generated by the flipping of Tyr-96, the molecular water network is further extended at the MBD$_{MBD4}$-DNA interface. For example, W1 forms a hydrogen-bonding network with other surrounding water molecules (Figs. 3*B* and 4), whereas its counterpart in the MBD$_{MeCP2}$-DNA complex, W1′ in Fig. 3*F*, is fixed by hydrogen bonds with the Tyr-123 and

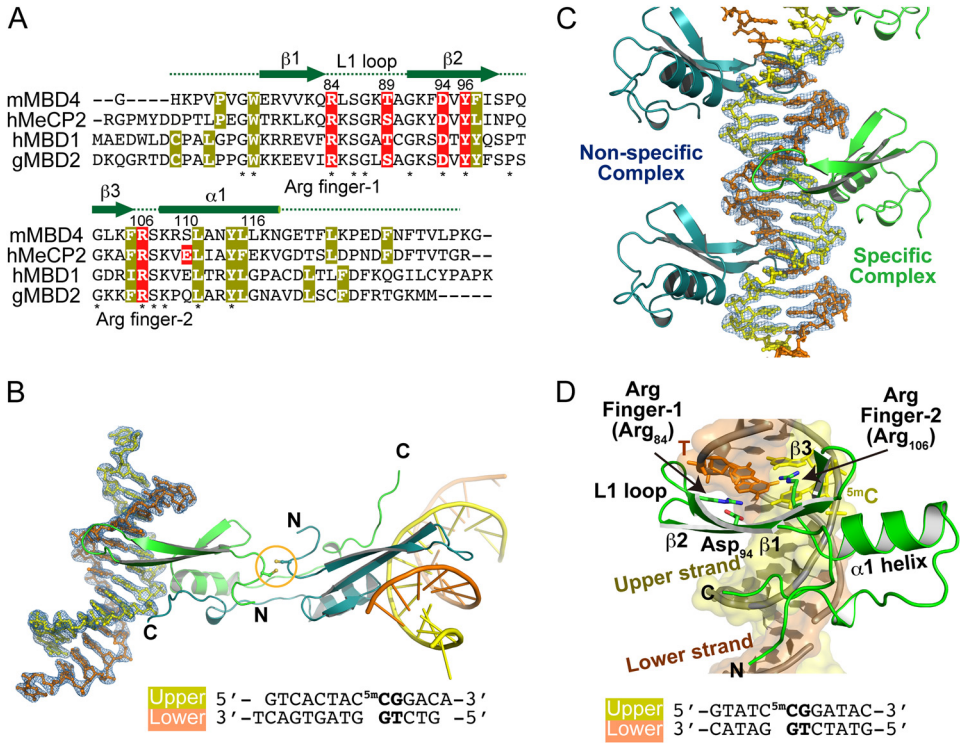

FIGURE 2. **Overall structure of MBD$_{MBD4}$ complexed with DNA containing ^{5m}CG/TG.** *A,* sequence alignment of structurally known MBD domains; mouse MBD4 (mMBD4, 69–136 amino acids), human MeCP2 (hMeCP2, 88–167 amino acids), human MBD1 (hMBD1, 1–75 amino acids), and chicken MBD2 (gMBD2, 3–72 amino acids). *Asterisks* indicate the conserved residues among the MBD proteins. The hydrophobic core residues are highlighted in *yellow*. The residues highlighted in *red* are involved in recognition of methylated CpG base pairs. *B* and *C*, overall structures of MBD$_{MBD4}$-^{5m}CG/TG complexes. *B*, $C222_1$ orthorhombic form; *C*, *P*1 triclinic form. The $2mF_o - DF_c$ electron density map for the DNA molecule contoured at 1.5 σ is shown in *blue*. The lower DNA strand containing a mismatched thymine is presented in *orange*, and the complementary upper strand is *yellow*. A disulfide bond linking two symmetry-related MBDs is indicated by an *orange circle* in *B*. In the *P*1 form, specific (*green*) and nonspecific (*blue*) MBD$_{MBD4}$-DNA complexes are evident (*C*). *D,* specific binding of MBD$_{MBD4}$ to ^{5m}CG/TG. The triclinic *P*1 form structure of MBD$_{MBD4}$ is shown as a *green ribbon* representation. Side chains of Arg-84, Arg-106, and Asp-94 are presented as stick models. The lower DNA strand containing a mismatched thymine is presented in *orange*, and the complementary upper strand is *yellow*. The ^{5m}CG/TG base pairs are shown as stick models. The DNA sequences are indicated below in *B* and *D*.

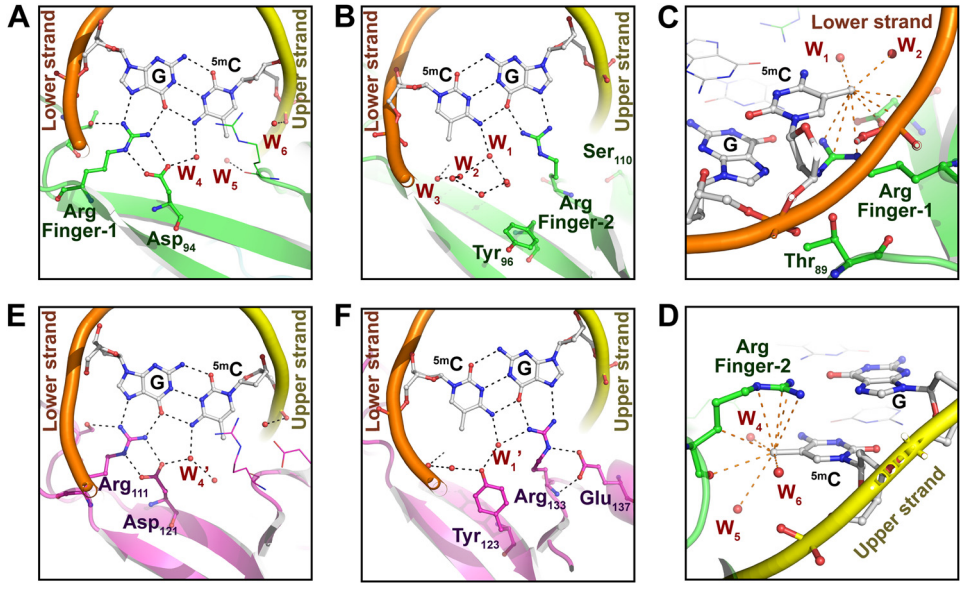

FIGURE 3. **Structural comparison of DNA binding surfaces of MBD$_{MBD4}$ and MBD$_{MeCP2}$.** *A* and *B,* recognition of the ^{5m}CG site by Arg finger-1 (*A*) or Arg finger-2 (*B*) of MBD$_{MBD4}$. Hydration water molecules are represented as *small red spheres*. DNA backbone structures of the upper and lower strands are depicted as *yellow* and *orange tubes*, respectively. *Black dotted lines* indicate hydrogen bonds (<3.2 Å). *C,* recognition of the methyl group of ^{5m}C in the lower strand by MBD$_{MBD4}$. *Orange dotted lines* indicate nonbonded contacts with the 5-methyl group of ^{5m}C (<4.2 Å). *D,* recognition of the 5-methyl group of the lower strand ^{5m}C by MBD$_{MBD4}$. The *red spheres* labeled as *W1*–*6* in *A*–*D* represent the water molecules in the coordinate file of the MBD$_{MBD4}$-^{5m}CG/^{5m}CG complex structure (PDB code 3VXX); *W1*, Wat-202 in chain C; *W2*, Wat-312 in chain A; *W3*, Wat-319 in chain A; *W4*, Wat301 in chain A; *W5*, Wat-214 in chain B; *W6*, Wat-202 in chain B. *E* and *F,* recognition of the ^{5m}CG site by Arg finger-1 (*E*) or Arg finger-2 (*F*) of MBD$_{MeCP2}$ (21). *Black dotted lines* indicate hydrogen bonds. W1′ and W4′ indicated in *E* and *F* correspond to the water molecules in the MeCP2-DNA complex structure (PDB code 3C2I), Wat-32 in chain B and Wat-183 in chain A, respectively.

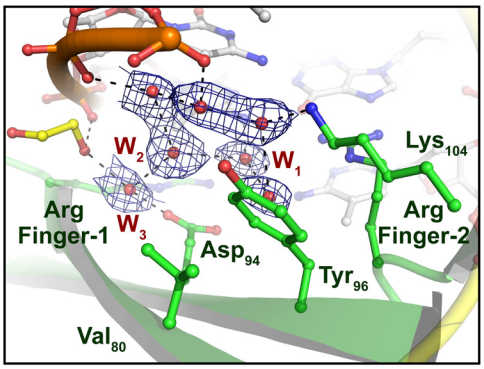

FIGURE 4. **The network structure of hydration water molecules within the interface of the MBD$_{MBD4}$-^{5m}CG/^{5m}CG complex.** The structure of the protein-DNA interface around Tyr-96 is magnified. The hydration water molecules around the lower strand ^{5m}C are shown as *red spheres* with a $2mF_o - DF_c$ electron density map contoured at 1.0 σ. The ethylene glycol (EG) molecule is shown as a yellow stick model.

Arg-133 residues of the protein. The hydrogen-bonding network within the DNA interface of MBD$_{MBD4}$ is also maintained through water-mediated interactions between the phosphate groups of the DNA backbone and the side chains of Asp-94 and Lys-104 (Fig. 4). Thus, the DNA interface of MBD$_{MBD4}$ contains more open space filled with ordered water molecules in comparison with other MBDs.

Recognition of ^{5m}CG/TG by the Flexible DNA Binding Surface of MBD$_{MBD4}$—The hydrogen bonding pattern of the T/G mismatched base pair in the MBD$_{MBD4}$-^{5m}CG/TG complex is identical to that observed in the crystal structure of a DNA oligomer with a T/G mismatch (PDB entry 113D) (34). The T/G mismatch still allows two hydrogen bonds to form between the bases, thus creating an overall shape similar to that in Watson-Crick base pairing. However, the mismatched thymine base is shifted 1–2 Å toward the major groove side of the DNA duplex (Fig. 5*A*). The base stacking interactions with neighboring pairs are unaffected by the mismatched pair (35), and the entire DNA binding mode common to MBDs is retained in the MBD$_{MBD4}$-^{5m}CG/TG complex.

The guanine base in the T/G mismatch is recognized by Arg finger-2 through a hydrogen-bonding pattern analogous to that observed in the MBD$_{MBD4}$-^{5m}CG/^{5m}CG complex (Fig. 5, *A* and *B*). However, in comparison with the ^{5m}CG/^{5m}CG complex, the side chain of Arg finger-2 is shifted by ~0.8 Å to form an additional hydrogen bond with the protruding carbonyl group at the 4th position of the thymine ring (Fig. 5*C*). Except for the movement of Arg finger-2, there are no significant differences between the protein structures in the ^{5m}CG/^{5m}CG and ^{5m}CG/TG complexes. The 5-methyl group of the thymine base is recognized via contacts with Arg finger-1, Asp-94, and water molecules as observed for the lower strand ^{5m}C recognition in the ^{5m}CG/^{5m}CG complex (Fig. 5*B*). It is important to note that the water molecules in the protein-DNA interface are rearranged by local conformational changes upon binding to ^{5m}CG/TG or ^{5m}CG/^{5m}CG (Fig. 5*C*).

The ^{5m}C base of ^{5m}CG/TG in the upper strand is also recognized by Arg finger-1 in the manner common to MBDs (Fig. 5*B*). Arg finger-2 retains the van der Waals contacts with the upper ^{5m}C base via its aliphatic moiety despite its movement.

In contrast to Arg finger-2 of MBD$_{MBD4}$, Arg finger-2 of MBD$_{MBD1}$ or MBD$_{MeCP2}$ is presumably incapable of recognizing the protruding mismatched base because its side chain is fixed by the interaction with conserved acidic residues (Fig. 3*F*) (20–22). Indeed, MBD$_{MBD1}$ exhibited significantly weaker binding to ^{5m}CG/TG compared with ^{5m}CG/^{5m}CG (Fig. 1; Table 3). Thus, the flexibility of Arg finger-2 provided by the lack of an intra-molecular lock appears to be indispensable for T/G mismatch recognition.

The Nonspecific DNA Binding Mode of MBD$_{MBD4}$—The nonspecific DNA binding mode of MBD4 is observed in the crystal structure of the triclinic form of the MBD4-^{5m}CG/TG complex (Fig. 2*C*). In the nonspecific complex, MBD$_{MBD4}$ also binds to DNA via the major groove side. The phosphate backbone recognition scheme by the α1 helix and L1 loop is essentially identical to that in the specific complex (Fig. 6*A*).

In the nonspecific complex, the dynamic movement of Arg finger-2 is of great interest; this movement takes place in the vacant space generated by the flipping of Tyr-96. Arg finger-2, which is directed toward the target base in the specific complexes, adopts a completely different conformation to form a hydrogen bond with an atom of the phosphate backbone (Fig. 6*B*) and, thereby reinforcing DNA duplex binding. The unique flexibility of Arg finger-2 in MBD4 presumably facilitates nonspecific DNA interaction, which implies a sliding mode prior to target recognition (Fig. 6*B*). In agreement with the structural observations, MBD$_{MBD4}$ exhibited more highly significant binding to nonmodified CpG than MBD$_{MBD1}$ in our electrophoretic mobility shift assay (data not shown).

The DNA Binding Surface of MBD4 Tolerates Binding to Oxidation and Deamination Products of ^{5m}C—The structural features of the protein-DNA interface suggest that MBD$_{MBD4}$ has the ability to bind to modifications that are more bulky than the methyl group at the 5th position of cytosine. We therefore examined the binding of MBD$_{MBD4}$ to a methylated CpG fragment containing ^{hm}C, ^{hm}U, ^{fo}C, or ^{ca}C (Fig. 7*A*). In a competitive EMSA, the nonlabeled ^{5m}CG/^{hm}CG, ^{5m}CG/^{hm}UG, and ^{5m}CG/^{fo}CG fragments competed with a ^{32}P-labeled ^{5m}CG/^{5m}CG duplex for binding to MBD$_{MBD4}$. The affinity of MBD$_{MBD4}$ for ^{5m}CG/^{hm}CG, ^{5m}CG/^{hm}UG, and ^{5m}CG/^{fo}CG was estimated to be ~2- or 3-fold weaker than its affinity for ^{5m}CG/^{5m}CG based on the data from the competitive EMSA and ITC binding assays (Fig. 7, *B* and *D*, and Table 3). However, the ^{5m}CG/^{ca}CG and ^{hm}CG/^{hm}CG fragments exhibited weaker binding to MBD$_{MBD4}$ than the other modified nucleotides (Fig. 7*B*). In contrast, MBD$_{MBD1}$ exhibited a tight specificity for ^{5m}CG/^{5m}CG (Fig. 7, *C* and *E*). The affinity of MBD$_{MBD1}$ for ^{5m}CG/^{hm}CG (K_D, 1.04 μM) was more than 10-fold weaker than that for ^{5m}CG/^{5m}CG (K_D, 72.5 nM) (Table 3). Combined with the structural data, these findings suggest that MBD$_{MBD4}$ is capable of binding to methylated CpG sequences that have undergone further asymmetric oxidative modification.

To achieve a better understanding of the structural basis of the versatile DNA binding ability of MBD$_{MBD4}$, we determined its crystal structure at 2.4-Å resolution when bound to a ^{5m}CG/^{hm}CG fragment (Table 1). Hydroxylation of the 5-methyl group of ^{5m}C does not perturb either the canonical hydrogen bonding pattern in the C/G base pair or the overall DNA binding mode

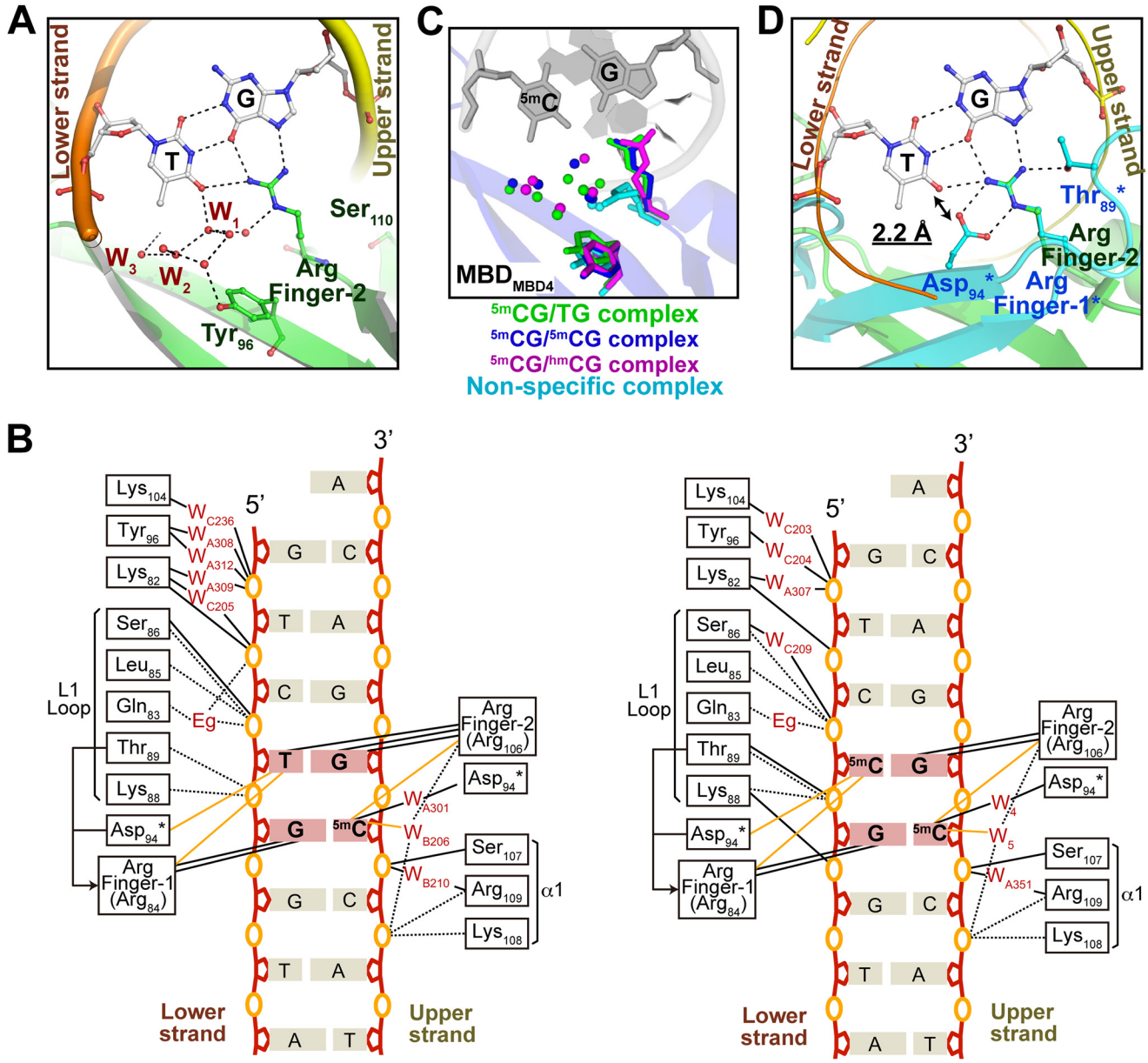

FIGURE 5. **Recognition of ^{5m}CG/TG by MBD$_{MBD4}$.** *A,* the structure of the T/G recognition site is shown in the same orientation as Fig. 3*B*. Hydration water molecules around the thymine base in the lower strand are shown as *red spheres*. *Black dotted lines* indicate hydrogen bonds (<3.2 Å). *W1–3* represent the water molecules in the coordinate file of the MBD$_{MBD4}$-^{5m}CG/TG complex structure (PDB code 3VXV); *W1*, Wat-202 in chain C; *W2*, Wat-206 in chain C; *W3*, Wat-318 in chain A. *B,* diagrams of the interactions between MBD$_{MBD4}$ and DNA in the ^{5m}CG/TG (*left panel*) and ^{5m}CG/^{5m}CG (*right panel*) complexes. Indirect protein-DNA interactions only mediated by a single water molecule are included in the diagram. *Solid* and *dotted black lines* indicate hydrogen bonds (<3.2 Å) donated from the main chain and side chain atoms of MBD$_{MBD4}$, respectively. *Solid orange lines* indicate nonbonded contacts with ^{5m}C or mismatched T (<4.2 Å). Asp-94 (*asterisk*) is shown on both sides of the schematic DNA drawing for convenience. *W* and *Eg* represent ordered water molecules and an ethylene glycol molecule, respectively. In the *right panel*, W4 and W5 correspond to Wat-301 of chain A and Wat-214 of chain B in the MBD$_{MBD4}$-^{5m}CG/^{5m}CG complex structure (PDB code 3VXX), respectively. *C,* comparison of the orientation of Arg finger-2 and the hydration water molecule network in the protein-DNA interfaces of the different complexes. Ribbon presentation represents the structure of the ^{5m}CG/^{5m}CG complex. *D,* model of MBD$_{MBD4}$ bound to the ^{5m}CG/TG sequence in the direction opposite to that observed in the crystal. The crystal structure of MBD$_{MBD4}$ is shown as a *green ribbon* representation with a *green stick* model of Arg finger-2. The model in the reverse binding mode is shown in *blue*. A guanidino group of Arg finger-1 in the model structure is overlaid onto that of Arg finger-2 in the crystal structure. The model suggests steric hindrance between the side chain of Asp-94 and the mismatched thymine base in the reverse binding mode.

of MBD$_{MBD4}$ (Fig. 8*A*). An unambiguous electron density for the hydroxyl group of ^{hm}C suggests a confined rotational movement of the 5-hydroxymethyl moiety against the pyrimidine ring (Fig. 8*A*); intriguingly, the hydroxyl group makes an intra-base hydrogen bond with the amino group at the 4th position in addition to a hydrogen bond with a water molecule at the DNA interface. The 5-hydroxymethyl moiety also donates CHO hydrogen bonds to the carbonyl of Asp-94 and the phosphate group of the DNA backbone, which show tetrahedral coordination around the methyl carbon at the 5th position (Fig. 8*B*). Thus, the positional preference of the hydroxyl group is ensured by the intra-base hydrogen bond and the tetrahedral configuration around the methyl carbon despite the close contacts with the neighboring base on the 5′ side. The flexible DNA interface of MBD$_{MBD4}$ is likely to have enough space to accommodate the ^{hm}U or ^{fo}C base as well as ^{hm}C. In contrast, the

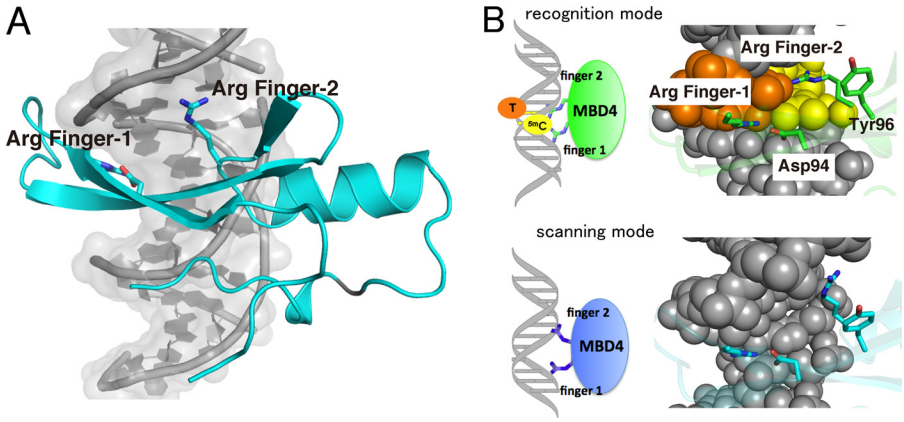

FIGURE 6. **Structure of the nonspecific DNA complex of MBD$_{MBD4}$.** *A,* structure of the nonspecific complex observed in the *P*1 form. DNA molecule is presented as a surface model. The protein residues are shown as *blue stick models*. *B,* schematic model of recognition and scanning modes of MBD$_{MBD4}$. Magnified views of the structure of Arg fingers in the nonspecific and the ^{5m}CG/TG complexes are shown in *right panels*.

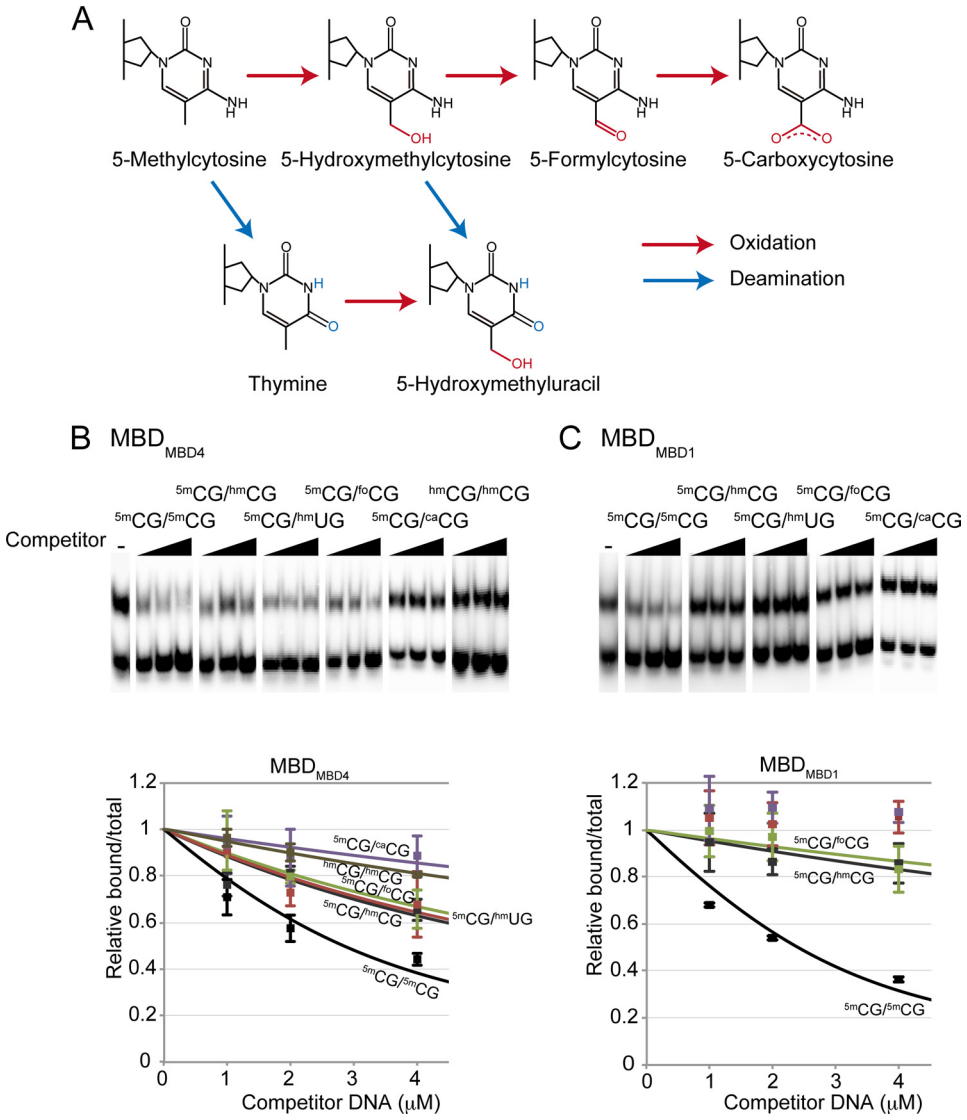

FIGURE 7. **Broad binding specificity of MBD$_{MBD4}$.** *A,* schematic representation of cytosine oxidation and deamination. *B* and *C,* DNA binding specificities of MBD$_{MBD4}$ (*B*) and MBD$_{MBD1}$ (*C*) analyzed by competitive electrophoretic mobility shift assay. Representative autoradiographic images of competitive assays with MBD$_{MBD4}$ and MBD$_{MBD1}$ are presented. The *left panel* of each section shows the control experiment in the absence of competitor. *D* and *E,* the relative values for each complex are plotted against the amount of competitor DNA. Each data point represents an average of three independent experiments using MBD$_{MBD4}$ or MBD$_{MBD1}$. Neither the ^{5m}CG/^{ca}CG nor the ^{5m}CG/^{hm}UG fragment exhibited competitive effects on the MBD$_{MBD1}$-^{5m}CG/^{5m}CG complex.

relatively low affinity of MBD$_{MBD4}$ for ^{5m}CG/^{ca}CG is presumably caused by electrostatic repulsion between the 5-carboxyl group of the base and the side chain carboxyl of Asp-94.

DISCUSSION

The crystal structures of MBD$_{MBD4}$ complexed with ^{5m}CG/TG, ^{5m}CG/^{5m}CG, and ^{5m}CG/^{hm}CG provide new insight into the structural mechanism of the versatility of base recognition by MBD4. The broad base specificity of MBD$_{MBD4}$ is implicated in heterochromatin localization and enzymatic activity of MBD4 associated with methylated DNA regions. In contrast to MBD$_{MBD1}$, MBD$_{MBD4}$ binds not only ^{5m}CG/^{5m}CG but also various modified pyrimidine rings including deamination and/or oxidation products of the ^{5m}C base, such as ^{5m}CG/TG, ^{5m}CG/^{hm}CG, ^{5m}CG/^{hm}UG, and ^{5m}CG/^{fo}CG, in a methylated CpG site. MBD$_{MBD4}$ shares an overall DNA recognition mode with other MBDs. The important role of the water molecules in target base recognition is highlighted by their conserved positions in the MBD$_{MBD4}$-^{5m}CG/^{5m}CG and MBD$_{MeCP2}$-^{5m}CG/^{5m}CG complexes (Fig. 3, *A*, *B*, *E*, and *F*) (21). However, local structural differences between MBD$_{MBD4}$ and other MBDs have a large impact on the DNA binding properties of MBD$_{MBD4}$. In particular, the structural features unique to MBD$_{MBD4}$ around the conserved Tyr-96 and the Arg finger-2 provide plasticity in the DNA binding surface and allow versatile base recognition (Fig. 5*C*). As a consequence of the flipped Tyr-96 side chain of MBD$_{MBD4}$, a more extensive water molecule network is established in its DNA interface compared with the MBD$_{MeCP2}$-DNA surface. Intriguingly, the hydration water molecules responsible for the base recognition (W1, W2, and W3) are maintained at appropriate positions through the solvent network rather than through interactions with protein residues (Fig. 4). A comparison of the water structures around the lower strand target bases (^{5m}C, mismatched T and ^{hm}C) highlighted the plasticity in the arrangement of the ordered water molecules in the DNA interface, in which the water-mediated hydrogen-bonding network of MBD$_{MBD4}$ is finely tuned to accommodate each of the modified bases.

Compared with the lower strand target base recognition, the interface with the upper strand ^{5m}C more strictly maintains the structural features conserved in other MBDs including the conformation of Arg finger-1, which is fixed by the aspartic acid and hydration water structure (Fig. 3, *A* and *E*). This structural feature obviously indicates that the symmetric oxidative modification of both ^{5m}C bases in the CpG sequence perturbs MBD$_{MBD4}$ binding. In fact, our DNA binding data combined with previously reported data demonstrate that neither MBD4 nor the other MBD proteins are capable of binding to the symmetrical ^{hm}CG/^{hm}CG site (Fig. 7, *B* and *C*) (36, 37). MBD$_{MBD4}$ does not make contact with bases other than the CpG sequence and are able to bind to the symmetric ^{5m}CG/^{5m}CG site equally in both directions as observed in the flipping motion of MBD$_{MBD1}$ on its target DNA (38). In contrast, the tight recognition of ^{5m}CG by the Arg finger-1 presumably prevents the flipping motion of MBD$_{MBD4}$ on asymmetric target sequences, such as ^{5m}CG/TG, ^{5m}CG/^{hm}CG, and ^{5m}CG/^{hm}UG (Fig. 5*D*).

Despite the broad spectrum of MBD$_{MBD4}$ binding targets, full-length MBD4 exhibits glycosylase activity only toward mismatched thymine and ^{hm}U bases (Fig. 9) (24, 39). The oxidative products of ^{5m}C, such as ^{hm}C, ^{fo}C, and ^{ca}C, are not susceptible to digestion by MBD4, whereas TDG excises ^{fo}C and ^{ca}C (7, 10). These findings indicate a partial functional redundancy and a possible functional difference between MBD4 and TDG (7, 10). The glycosylase domain itself exhibits the substrate specificity for T/G or ^{hm}U/G mismatched bases regardless of the methylation status of the adjacent C/G base pair (23, 25). Therefore, the DNA binding of MBD$_{MBD4}$ is presumably a prerequisite for the intrinsic glycosylase activity of MBD4 toward the mismatched bases generated in methylated CpG sites. Addition-

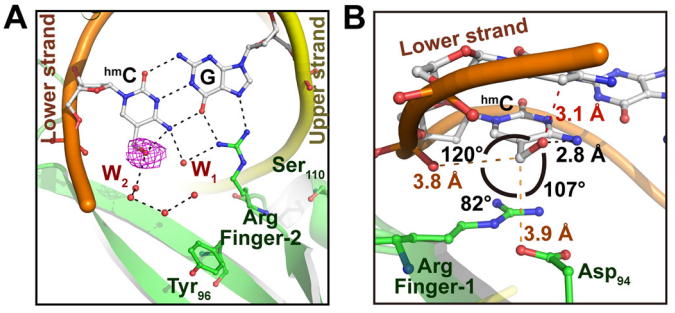

FIGURE 8. **Structure of MBD$_{MBD4}$ bound to ^{5m}CG/^{hm}CG.** *A*, structure of the DNA interface in the MBD$_{MBD4}$-^{5m}CG/^{hm}CG complex. The structure of the ^{hm}CG binding site is shown in the same orientation as Fig. 3*B*. The $mF_o - DF_c$ simulated annealing omit map (>3.0 σ) for the hydroxyl group of the ^{hm}C base is shown as *magenta mesh*. Water molecules are represented as *small red spheres*. *Black dotted lines* indicate hydrogen bonds (<3.2 Å). *W1* and *W2* represent the water molecules in the PDB file of the MBD$_{MBD4}$-^{5m}CG/^{hm}CG complex structure (PDB code 3VYB); *W1*, Wat-102 in chain C; *W2*, Wat-106 in chain C. *B*, the tetrahedral configuration around the carbon atom in the 5-hydroxymethyl group. The *black*, *orange*, and *red dotted lines* represent a hydrogen bond, a CHO hydrogen bond, and an unfavorable close contact, respectively.

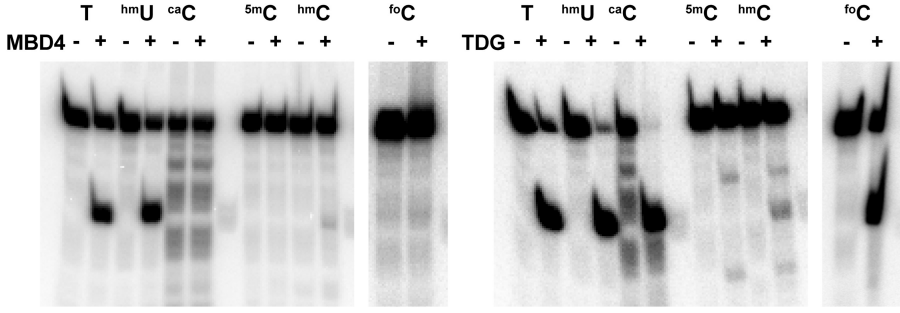

FIGURE 9. **Glycosylase activities of MBD4 and TDG.** The glycosylase activity of the full-length MBD4 protein for mismatched, deamination, and/or oxidation products in the context of the ^{5m}CG/^{5m}CG sequence was assessed by NaOH cleavage of the resulting apyrimidinic site. We observed a significant digestion band for the strand containing either T or ^{hm}U in a mismatched wobble base. ^{5m}C, ^{hm}C, ^{fo}C, and ^{ca}C bases, each of which forms canonical Watson-Crick base pairs, were not removed by MBD4, whereas, human TDG exhibited activity toward ^{fo}C and ^{ca}C in addition to T and ^{hm}U bases.

ally, isolated MBD_{MBD4} has been shown to inhibit the catalytic activity of the glycosylase domain toward a single $^{5m}CG/TG$ site *in vitro* (40), suggesting that the DNA substrate is transferred from MBD_{MBD4} to the glycosylase domain only in full-length MBD4. The unidirectional binding of MBD_{MBD4} to the $^{5m}CG/TG$ or $^{5m}CG/^{hm}UG$ site could facilitate its synergetic action with the C-terminal glycosylase domain in DNA mismatch repair processes. It remains unclear whether the binding of MBD_{MBD4} to ^{5m}C, ^{hm}C, or ^{fo}C targets the glycosylase domain to neighboring $^{5m}CG/TG$ or $^{5m}CG/^{hm}UG$ sites.

Intriguingly, the active DNA demethylation of the *p15^{ink4b}* tumor suppressor gene triggered by the TGF-β/SMAD signaling pathway is accompanied by the accumulation of ^{hm}C bases, MBD4, TDG, and downstream base excision repair proteins (19). The versatile base recognition ability of MBD_{MBD4} demonstrated in our study may contribute to the stimuli-dependent accumulation of MBD4 at hydroxymethylated regions, which leads to erasure of DNA methylation marks. Further investigation of MBD4 protein complexes colocalized to ^{hm}C-rich regions will be crucial for fully understanding the functional roles of MBD4 in DNA demethylation pathways. Furthermore, recent studies have indicated that the ^{hm}C, ^{fo}C, and ^{ca}C bases have long lifetimes during preimplantation development (41, 42); thus they may function as *bona fide* epigenetic marks antagonistic to ^{5m}C bases *in vivo*. MBD4 may recognize these bases independently of its glycosylase activity and act as a mediator via its multifunctional capabilities, although further investigation is necessary to fully understand the role of MBD4 in the biology of oxidized cytosine bases.

Acknowledgments—We thank Drs. N. Matsugaki, N. Igarashi, and Y. Yamada and Prof. S. Wakatsuki for data collection at Photo Factory, Tsukuba. The synchrotron radiation experiments at the BL38B1 were performed with the approval of the Japan Synchrotron Radiation Research Institute (JASRI) (Proposal No. 2010B1059).

REFERENCES

1. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6–21
2. Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3,** 662–673
3. Robertson, K. D., and Wolffe, A. P. (2000) DNA methylation in health and disease. *Nat. Rev. Genet.* **1,** 11–19
4. Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R. Z., Carmouche, R. P., Ibberson, D., Barath, P., Demay, F., Reid, G., Benes, V., Jeltsch, A., Gannon, F., and Salbert, G. (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452,** 45–50
5. Kangaspeska, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R. P., Benes, V., Gannon, F., and Reid, G. (2008) Transient cyclical methylation of promoter DNA. *Nature* **452,** 112–115
6. Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., Abramowitz, L. K., Bartolomei, M. S., Rambow, F., Bassi, M. R., Bruno, T., Fanciulli, M., Renner, C., Klein-Szanto, A. J., Matsumoto, Y., Kobi, D., Davidson, I., Alberti, C., Larue, L., and Bellacosa, A. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146,** 67–79
7. He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307
8. Wu, S. C., and Zhang, Y. (2010) Active DNA demethylation. Many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11,** 607–620
9. Guo, J. U., Su, Y., Zhong, C., Ming, G. L., and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145,** 423–434
10. Maiti, A., and Drohat, A. C. (2011) Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine. Potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286,** 35334–35338
11. Bhutani, N., Brady, J. J., Damian, M., Sacco, A., Corbel, S. Y., and Blau, H. M. (2010) Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463,** 1042–1047
12. Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., Jacobsen, S. E., and Reik, W. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463,** 1101–1105
13. Rai, K., Huggins, I. J., James, S. R., Karpf, A. R., Jones, D. A., and Cairns, B. R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135,** 1201–1212
14. Bogdanović, O., and Veenstra, G. J. (2009) DNA methylation and methyl-CpG binding proteins. Developmental requirements and function. *Chromosoma* **118,** 549–565
15. Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005) The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16^{INK4a} and *hMLH1* genes. *Mol. Cell. Biol.* **25,** 4388–4396
16. Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. (1999) The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* **401,** 301–304
17. Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R., and Bird, A. (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297,** 403–405
18. Riccio, A., Aaltonen, L. A., Godwin, A. K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D. E., Ruggeri, B. A., Klein-Szanto, A. J., Testa, J. R., Neri, G., and Bellacosa, A. (1999) The DNA repair gene *MBD4* (*MED1*) is mutated in human carcinomas with microsatellite instability. *Nat. Genet.* **23,** 266–268
19. Thillainadesan, G., Chitilian, J. M., Isovic, M., Ablack, J. N., Mymryk, J. S., Tini, M., and Torchia, J. (2012) TGF-β-dependent active demethylation and expression of the p15^{ink4b} tumor suppressor are impaired by the ZNF217/CoREST complex. *Mol. Cell*, **46**, 636–649
20. Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105,** 487–497
21. Ho, K. L., McNae, I. W., Schmiedeberg, L., Klose, R. J., Bird, A. P., and Walkinshaw, M. D. (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell* **29,** 525–531
22. Scarsdale, J. N., Webb, H. D., Ginder, G. D., and Williams, D. C. (2011) Solution structure and dynamic analysis of chicken MBD2 methyl binding domain bound to a target-methylated DNA sequence. *Nucleic Acids Res.* **39,** 6741–6752
23. Manvilla, B. A., Maiti, A., Begley, M. C., Toth, E. A., and Drohat, A. C. (2012) Crystal structure of human methyl-binding domain IV glycosylase bound to abasic DNA. *J. Mol. Biol.* **420,** 164–175
24. Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X., and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.*, **40**, 4841–4849
25. Moréra, S., Grin, I., Vigouroux, A., Couvé, S., Henriot, V., Saparbaev, M., and Ishchenko, A. A. (2012) Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA. *Nucleic Acids Res.*, **40**, 9917–9926
26. Doublié, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276,** 523–530
27. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol. A* **276,** 307–326
28. Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 849–861

29. Terwilliger, T. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56,** 965–972
30. Emsley, P., and Cowtan, K. (2004) COOT. Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132
31. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX. A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 213–221
32. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Cα geometry. π, ψ and Cβ deviation. *Proteins* **50,** 437–450
33. Morrison, J. F. (1969) Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim. Biophys. Acta* **185,** 269–286
34. Hunter, W. N., Brown, T., Kneale, G., Anand, N. N., Rabinovich, D., and Kennard, O. (1987) The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.* **262,** 9962–9970
35. Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X. J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C. S., Westhof, E., Wolberger, C., and Berman, H. M. (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **313,** 229–237
36. Jin, S. G., Kadam, S., and Pfeifer, G. P. (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* **38,** e125
37. Valinluck, V., Tsai, H. H., Rogstad, D. K., Burdzy, A., Bird, A., and Sowers, L. C. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **32,** 4100–4108
38. Inomata, K., Ohki, I., Tochio, H., Fujiwara, K., Hiroaki, H., and Shirakawa, M. (2008) Kinetic and thermodynamic evidence for flipping of a methyl-CpG binding domain on methylated DNA. *Biochemistry* **47,** 3266–3271
39. Liu, P., Burdzy, A., and Sowers, L. C. (2003) Repair of the mutagenic DNA oxidation product, 5-formyluracil. *DNA Repair* **2,** 199–210
40. Aziz, M. A., Schupp, J. E., and Kinsella, T. J. (2009) Modulation of the activity of methyl binding domain protein 4 (MBD4/MED1) whereas processing iododeoxyuridine generated DNA mispairs. *Cancer Biol. Ther.* **8,** 1156–1163
41. Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011) Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21,** 1670–1676
42. Inoue, A., and Zhang, Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334,** 194
43. DeLano, W. L. (2010) *The PyMOL Molecular Graphics System*, Schrödinger, LLC, New York